Reclassement d'images par le contenu

Georges Quénot, Franck Thollard

Reclassement d'images par le contenu

Georges Quénot — Franck Thollard

UJF-Grenoble 1 / UPMF-Grenoble 2 / Grenoble INP / CNRS, LIG UMR 5217, Grenoble, F-38041, France
{Prénom.Nom}@imag.fr

RÉSUMÉ. *Cet article présente une méthode permettant de reclasser les images fournies par un moteur de recherche par mots-clés à l'échelle du web et à l'état de l'art. Cette méthode utilise le contenu visuel des images et elle est basée sur l'idée que les images pertinentes doivent être semblables entre elles et que les images non pertinentes doivent être différentes entre elle et des images pertinentes. Cette idée a été implémentée en classant les images en fonction de la distance moyenne de celles-ci avec leurs plus proches voisines. Le reclassement seul selon cette méthode ne fait pas mieux que le classement original du système de recherche mais, combiné à celui-ci, il permet un gain en performance relatif d'environ 3% en termes de précision moyenne et d'environ 7% si l'on considère le complément à 1 de la précision moyenne. Ce gain est statistiquement significatif et il est apporté à un système déjà très performant pour l'exploitation de l'ensemble des informations textuelles associées aux images.*

ABSTRACT. *This article presents a method for re-ranking images retrieved by a web scale and state of the art search engine using key words for entering queries. This method uses the visual content of the images and it is based on the idea that the relevant images should be similar to each other while the non relevant images should be different from each other and from relevant images. This idea has been implemented by ranking the images according to their average distances to their nearest neighbors. The re-ranking performed with this method only does not perform better than the original ranking but, when combined to it, it yields a relative performance gain of about 3% in term of mean average precision and of about 7% if the complement to 1 of the average precision is considered. This gain is statistically significant and it is brought to a system which is already very effective for exploiting all the textual information available with the images.*

MOTS-CLÉS : *Recherche d'images, reclassement, contenu visuel.*

KEYWORDS: *Image retrieval, re-ranking, visual content.*

CORIA 2012, pp. 203–214, Bordeaux, 21-23 mars 2012

1. Introduction

La recherche d'information multimédia fait face au problème du fossé sémantique que l'on rencontre quand on cherche à interpréter une matrice de nombres conformément à ce que ferait un humain. La recherche d'images, dans le cadre du Web, peut se décomposer en différentes catégories : la recherche d'images à partir d'images, la recherche d'images à partir de caractéristiques visuelles, ou la recherche d'images à partir d'une requête textuelle classique.

Dans le premier cas, la requête étant une image, on va chercher à retrouver des images proches de l'image requête. Certains systèmes[1] suivent cette approche. Si l'image requête est disponible sur plusieurs sites, le système va retourner, en tête de liste, différentes versions de la même image ainsi que les sites sur lesquels elles ont été trouvées. Par ailleurs, `Google image` permet de glisser/déposer une image dans la barre de requête. Il semble que l'information de couleur soit très importante dans la recherche par le contenu proposée par `Google`.

La seconde approche semble intéressante du point de vue développement des systèmes. Malheureusement, l'interface avec l'utilisateur est difficile à mettre en place car, bien que l'information de couleur soit relativement facile à fournir, l'information de texture ou de contour est plus difficile à transmettre par l'utilisateur. On pourra cependant citer le système QBIC (Flickner *et al.,* 1995) qui suit cette piste.

Dans la troisième approche, l'utilisateur fournit une requête textuelle pour laquelle le système produit une liste d'images. Cette stratégie semble être suivie par les moteurs de recherches généralistes tels que `Google`, `Yahoo !`, `picSearch`, `exalead`, etc. Bien que les moteurs généralistes dévoilent peu leurs stratégies, on notera que rechercher la même information (par exemple des photos de pluie) via des langues différentes (par exemple « rain » et « pluie ») produit des résultats très différents. On remarquera de plus que, outre les images retournées, l'information textuelle mise en exergue dans les résultats est très liée aux termes de la requête. Pour exemple, la requête « rain » produit des résultats liés à l'acteur-chanteur coréen « Bi-rain ». Ces résultats n'apparaissent pas comme résultat de la requête « pluie ». En revanche, la recherche « pluie » retourne aussi l'affiche du film « les passagers de la pluie ». Ceci nous indique que le poids du texte dans le traitement de la requête est prépondérant. Comme nous le verrons par la suite, cette stratégie est pertinente dès lors que l'on s'intéresse à l'aspect précision.

Nous proposons ici une approche dans laquelle un moteur de recherche généraliste est d'abord utilisé pour produire, à partir de mots-clés, une liste ordonnée d'images candidates. Cette recherche est `a priori` effectuée indépendamment de leur contenu visuel en se basant sur leur URL, leurs métadonnées et/ou sur le texte environnant celles-ci dans les documents dans lesquelles elles apparaissent. Dans une seconde étape, nous effectuons un classement alternatif basé uniquement sur leur contenu visuel. Enfin, nous fusionnons le classement original avec le classement alternatif pour produire un classement final prenant en compte les deux modalités, textuelles et vi-

1. On pourra citer par exemple `TinEye` (`http://www.tineye.com/`) et `Google image`.

suelles. Le classement alternatif est basé sur l'idée que les images pertinentes doivent se ressembler entre elles tandis que les images non pertinentes doivent être très diverses. Cette approche a l'avantage de garder l'interface textuelle, très pratique pour l'utilisateur, et d'utiliser tout de même le contenu visuel pour améliorer la qualité du résultat. On peut aussi l'appliquer comme un post-traitement simple, léger et local sur la sortie d'un système généraliste à l'état de l'art et à l'échelle du web.

Le suite de cet article est organisée comme suit : dans la section 2, nous présentons les travaux connexes à notre travail ; dans la section 3, nous présentons notre approche incluant pour une fonction de classement alternative basée sur le contenu et une méthode de fusion entre classements pour produire le classement final ; dans la section 4, nous présentons les résultats que nous avons obtenus sur une collection de 200 requêtes résolues collectées par le laboratoire GREYC dans le cadre du programme Quaero.

2. Travaux connexes

Notre travail porte sur le reclassement d'images obtenues par une recherche textuelle en fonction de leurs caractéristiques visuelles. Il est lié de ce point de vue à la recherche d'images à partir d'une ou plusieurs images requête. Celle-ci s'effectue en général en deux ou trois étapes :

– construction d'une représentation dite de bas niveau de l'image ou extraction de descripteurs ;

– utilisation d'une mesure de similarité entre ces représentations ou ces descripteurs supposée représentative de la similarité visuelle ou sémantique entre les images ;

– dans le cas où plusieurs descripteurs et/ou plusieurs mesures de similarité peuvent être utilisés, une étape de fusion peut être considérée.

Les images candidates sont finalement classées en fonction de leurs similarités ou de leurs distances avec la ou les images requêtes. Nous utiliserons plutôt ici des mesures de distance (à minimiser) que de similarité (à maximiser).

2.1. *Descripteurs*

Pour représenter le contenu des images, nous avons retenu des descripteurs de bas niveau classiques afin de capturer différents type d'information :

– couleur : histogrammes ou des sacs de couleurs (Wengert *et al.,* 2011) ;

– texture : transformées de `Gabor` (Zhang *et al.,* 2000) ; représentations à base de motifs binaires locaux (Mu *et al.,* 2008, Chechik *et al.,* 2010) en anglais LPB) ;

– points d'intérêts : par exemple les SIFT (Lowe, 2004) ou des variantes dont le calcul est plus performant (Bay *et al.,* 2006) ou la représentation plus compacte (Ke *et al.,* 2004, Jégou *et al.,* 2010).

2.2. *Distances*

Plusieurs types de distances ou de fonction qui ne sont pas rigoureusement des distances mais qui peuvent être utilisées pour évaluer la similarité entre images peuvent être considérées. Certaines sont particulièrement adaptées lorsque les descripteurs considérés sont des histogrammes comme la distance dite du χ^2. Nous considérons dans ce travail les trois distances suivantes, définies entre deux vecteurs $(x_i)_{1 \leq i \leq n}$ et $(y_i)_{1 \leq i \leq n}$ par :

– distance euclidienne : $d(x,y) = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2}$

– distance du χ^2 : $d(x,y) = \sum_{i=1}^{i=n} \frac{(x_i - y_i)^2}{x_i + y_i}$

– « distance » angulaire : $d(x,y) = \arccos\left(\frac{\sum_{i=1}^{i=n} x_i y_i}{\sqrt{\sum_{i=1}^{i=n} x_i^2}\sqrt{\sum_{i=1}^{i=n} y_i^2}}\right)$

Les deux dernières ne sont appropriées que pour les cas où tous les x_i et y_i sont positifs ou nuls, ce qui est le cas par exemple pour les histogrammes.

2.3. *Fusion*

Dans le cas où plusieurs descripteurs et/ou plusieurs mesures de similarité peuvent être utilisés, deux types de fusion peuvent être considérés selon que la fusion est effectuée avant ou après l'étape de mesure de similarité (ou de calcul de distance) :

– fusion précoce : ce sont les différents descripteurs qui sont fusionnés, en général par simple concaténation des vecteurs qui les constituent en un vecteur global unique ; pour qu'une telle fusion soit efficace, il est en général nécessaire de normaliser les différents descripteurs de façon à ce qu'ils aient tous une dynamique comparable, par exemple en changeant leur échelle de telle sorte que la distance moyenne entre deux échantillons sur une collection de développement soit constante.

– fusion tardive : ce sont les distances entre les différents descripteurs qui sont fusionnées, en général par moyenne arithmétique ou par combinaison linéaire ; là encore, pour qu'une telle fusion soit efficace, il est en général nécessaire de normaliser les différentes distances de façon à ce qu'ils aient tous une dynamique comparable.

3. Reclassement des images

3.1. *Fonction de classement basée sur le contenu*

Notre objectif est de reclasser les images d'une liste retournée par le moteur généraliste de recherche par mots-clés. Notre travail repose sur une intuition ou une hypothèse que l'on peut formuler des manières suivantes :

– le sous-ensemble des images pertinentes doit être homogène et le sous-ensemble des images non pertinentes doit être hétérogène et éloigné du sous-ensemble des images pertinentes ;

– les représentations des images pertinentes doivent être regroupées alors que les images non pertinentes doivent être dispersées ;

– les images pertinentes doivent avoir plus de similarités entre elles que les images non pertinentes entre elles ou avec les images pertinentes ;

– une image pertinente doit avoir une distance moyenne à ses voisines les plus proches plus faible qu'une image non pertinente.

Ces formulations ne sont pas tout à fait équivalentes ni d'ailleurs très précises mais elles correspondent plus ou moins à la même idée. Il n'est pas évident non plus a priori que cette idée soit correcte ni qu'elle puisse aider à mieux reclasser les images. Le but du travail présenté ici est d'évaluer dans quelle mesure elle peut effectivement aider pour cette tâche en partant de la dernière formulation et en commençant par préciser ce que l'on peut entendre par « distance moyenne à ses voisines les plus proches ». Une fois ceci défini, on peut construire un score pour classer les images à partir de là : le score d'une image donnée sera d'autant plus grand que la distance moyenne à ses voisins est petite.

Notons que notre approche n'est pas adaptée si des groupes d'images presque identiques existent dans notre collection. En effet, dans ce cas, notre système va tendre à classer les images par groupe d'images identiques. Bien que cela améliore probablement les résultats du point de vue de notre métrique, cela peut ne pas correspondre aux attentes de l'utilisateur. Ce problème a été traité dans des campagnes d'évaluation comme imageClef 2009 où la qualité de la diversité était prise en compte dans le classement des systèmes. Dans le cas de `Google`, les listes retournées contiennent assez peu d'images similaires, ce problème n'est donc pas forcément critique.

Un descripteur x et une distance entre descripteurs d étant choisis, on peut définir une distance moyenne d'un échantillon x_i à ses k voisins les plus proches par :

$$d_m(x_i) = \frac{\sum_{j=1}^{j=k} d(x_i, x_{n(i,j)})}{k}$$

où $n(i,j)$ est l'indice du $j^{\text{ème}}$ voisin le plus proche de l'échantillon d'indice i. Les voisins considérés comme plus proches le sont au sens de la même distance d.

On peut également considérer des variantes dans lesquelles une fonction puissance est appliquée à la distance entre échantillons de façon à adapter au mieux la dynamique de la distribution de cette distance et une fonction de pondération f de façon à réduire l'éventuel effet de fenêtrage lié au fait qu'on arrête brutalement de prendre en compte les voisins à une certaine profondeur :

$$d_m(x_i)^\alpha = \frac{\sum_{j=1}^{j=k} f(k) d(x_i, x_{n(i,j)})^\alpha}{\sum_{j=1}^{j=k} f(k)}$$

k, α et f sont des paramètres définissant une variante donnée de la fonction d_m. Une valeur optimale pour ces paramètres peut être déterminée par validation sur un ensemble de développement contenant des requêtes résolues. Le choix d'un descripteur et d'une distance entre descripteurs sont également des facteurs déterminant la « distance moyenne d'une image à ses voisines les plus proches ». Les images dans la liste retournée par le moteur de recherche généraliste sont finalement reclassées selon les valeurs croissante de la fonction d_m appliquée au descripteur x de ces images.

3.2. *Fusion de classements*

Les systèmes de recherche d'images par le contenu ou de classification d'images font souvent appel à plusieurs descripteurs, chacun capturant un aspect particulier du contenu visuel comme la distribution de couleurs, la distribution de textures ou la sélection de points caractéristiques. La méthode de reclassement décrite ci-dessus peut être appliquée sur plusieurs descripteurs séparément, les résultats étant ensuite fusionnés (fusion tardive). Alternativement, on peut également fusionner directement les descripteurs par concaténation (fusion précoce) ou les distances entre les descripteurs avant calcul de la distance moyenne (fusion intermédiaire). Selon le cas, il convient de normaliser de manière appropriée, soit les descripteurs (x_i), soit les distances entre eux ($d(x_i, x_j)$), soit les distances moyennes aux voisins les plus proches ($d_m(x_i)$).

Dans le cas présent, il est également possible de fusionner le classement initial produit par le moteur de recherche et le classement produit par la méthode de la distance aux plus proches voisins. Par contre, les moteurs de recherche généralistes fournissent un ordre pour les résultats qu'ils retournent mais ils ne fournissent en général pas de score associé que l'on pourrait normaliser pour le combiner au score obtenu avec la méthode de reclassement proposée.

Dans les cas où seulement des classements sont disponibles, un moyen de les fusionner est d'utiliser des méthodes à base de rang qui reviennent à construire un score pour chaque image à partir de son rang selon le classement considéré. Dans la méthode du comptage de Borda normalisée, on prend la valeur $(n - k)/(n - 1)$ comme score pour l'échantillon classé $k^{\text{ème}}$ parmi n. Cette méthode a l'inconvénient de donner le même poids à une différence entre deux échantillons successifs quels que soient leur position dans la liste alors que les différences en tête de liste doivent avoir un impact plus important que les différences en fin de liste. Pour tenir compte de cet effet, nous proposons d'appliquer une fonction puissance sur le score précédent. La fusion entre deux classements se fait alors par combinaison linéaire des scores.

Une fusion selon cette méthode dépend de trois paramètres : les exposants appliqués sur les deux classements à fusionner et le paramètre définissant leurs poids relatifs dans la combinaison linéaire. Comme précédemment une valeur optimale pour ces paramètres peut être déterminée par validation sur un ensemble de développement contenant des requêtes résolues. Nous utilisons cette même méthode de fusion tardive

pour fusionner les classements obtenus avec différents descripteurs et pour fusionner le classement par le contenu et le classement initial du moteur de recherche.

4. Expérimentations

4.1. *La collection*

Nous avons évalué notre approche en utilisant un corpus de requêtes résolues produit par le laboratoire GREYC dans le cadre du programme Quaero. Le corpus a été constitué en collectant le résultat de requêtes mono-termes réalisées auprès du moteur de recherche `Google image`. Les requêtes ont été décomposées sémantiquement en deux niveaux : un premier niveau appelé catégorie (constitué de 14 éléments) et un second niveau appelé concept (constitué de 519 éléments). Les catégories sont les suivantes : `animal`, `cartoon events`, `general`, `holiday`, `logo`, `object_personal`, `object_public`, `people`, `plant`, `scene_human`, `scene_natural`, `scene_tour`, `sports`. Les concepts sont des sous-catégories qui peuvent être des objets comme par exemple `binocular`, `violins`, `bicycles`, ou des concepts plus abstraits comme `valentine_day` ou `car_crash`.

Au final, la collection contient 187.029 images divisées en 14 catégories divisées en 200 concepts (100 concepts pour le développement, 100 concepts pour l'ensemble de test). Dans l'ensemble de développement, le nombre de concepts pour une même catégorie varie de 2 pour `holiday` à 65 pour `object_personal`. Le nombre d'images par concept est en moyenne de 935 avec une faible variance (~ 30). Toutes les images ont été manuellement annotées comme étant pertinentes ou non pour la requête ayant servi à les collecter.

4.2. *Mesure de qualité*

Dans notre cadre, les requêtes sont constituées d'un seul mot (les concepts), les systèmes renvoient, pour chaque concept une liste triée de résultats, une erreur en tête de liste pénalise plus qu'une erreur en fin de liste. Une mesure appropriée dans ce contexte est la précision moyenne (AP[2]). Dans la mesure où nous disposons de plusieurs concepts, la mesure de qualité finale sera la MAP[3] calculée en utilisant le programme `trec_eval`.

4.3. *Évaluation*

Nous comparons différentes versions de notre méthode entre elles et avec deux références : le classement produit par un système classant les images de manière aléa-

2. Le terme AP vient de l'anglo-saxon : Average Precision.
3. Le terme MAP vient de l'anglo-saxon : Mean Average Precision.

toire et le classement fourni par système initial de recherche par mots-clés. Nous présentons les résultats obtenus sur l'ensemble de développement sur lequel nous avons optimisé tous les paramètres, soit pour le reclassement initial, soit pour la fusion des classements, et sur l'ensemble de test sur lequel nous appliquons notre méthode avec les paramètres réglés sur l'ensemble de développement.

En plus des deux références, nous avons effectué quatre expérimentations distinctes :

– un reclassement simple en utilisant un descripteur de type « sac de mots visuels » sur des descripteurs SIFT couleur (opponent SIFT) calculés avec un programme de Koen van de Sande (van de Sande *et al.,* 2010). Nous avons retenu une variante avec échantillonnage dense et un histogramme flou à 1000 composantes ;

– un reclassement simple en utilisant un descripteur qui est une fusion précoce normalisé d'un histogramme RGB $4 \times 4 \times 4$ pour représenter la distribution de couleurs et d'une transformée de `Gabor` à 8 orientations et 5 échelles pour représenter la distribution de textures ;

– une fusion tardive des résultats des deux méthodes de reclassements simples basées sur le contenu visuel ;

– une fusion tardive du reclassement précédent et du classement initial fourni par le système de recherche par mots-clés.

Avant le calcul des descripteurs, toutes les images ont été redimensionnées de telle sorte que leur plus petite dimension fasse au moins 200 pixels et que leur plus grande dimension fasse au plus 400 pixels en conservant le rapport d'aspect chaque fois que cela est compatible avec ces deux contraintes. Ce redimensionnement a été effectué de façon à rendre plus comparables les descripteurs extraits, ceux de texture et de points d'intérêt étant peu robustes aux changements d'échelle et les images collectées ayant des tailles très variables.

Pour chacun des descripteurs, les paramètres k et α ont été optimisés sur l'ensemble de développement et pour chacune des trois distances considérées. Seule la distance donnant la meilleure performance une fois les paramètres optimisés a été retenue. Des essais avec une fonction f de lissage de la fenêtre ont été effectués mais ils n'ont pas permis d'améliorer la performance et aucune fonction f n'a été utilisée dans les résultats présentés.

Pour les fusions, la version tardive seulement a été considérée et elle a été faite en utilisant la méthode basée sur les rangs uniquement, même pour la fusion entre les descripteurs pour lesquels des scores étaient disponibles. Les deux exposants utilisés pour la normalisation et le paramètre de combinaison linéaire ont là aussi été optimisés sur l'ensemble de développement.

Dans tous les cas les paramètres sont optimisés en maximisant la précision moyenne sur l'ensemble de développement. Lorsque plusieurs paramètres sont à optimiser conjointement, l'optimisation globale est faite alternativement sur chaque pa-

ramètre en suivant à chaque fois le minimum local. Les valeurs précises de ces paramètres dépendent du type de descripteur.

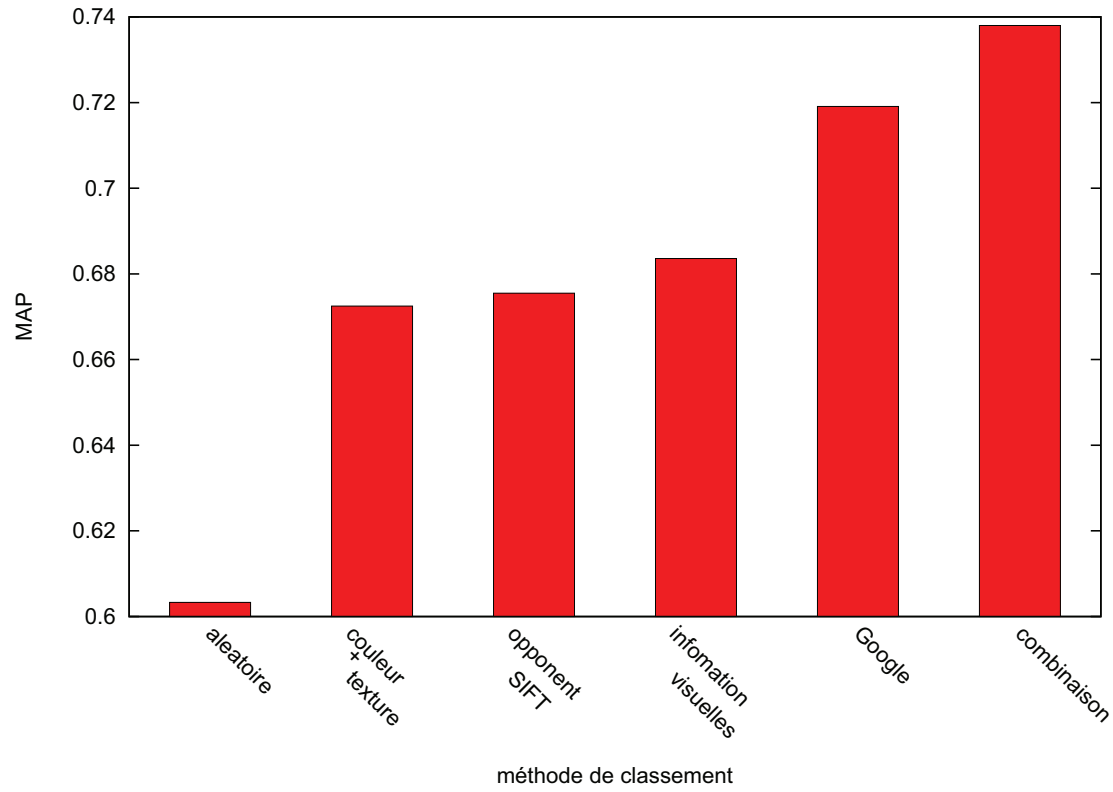

Figure 1. *Performances obtenues par les différentes méthodes sur l'ensemble d'apprentissage*

Les figures 1 et 2 montrent les résultats (MAP) obtenus lors de nos expérimentations ainsi que pour les deux références. Nous pouvons faire les observations suivantes :

– les valeurs absolues entre les collections de développement et de test sont comparables avec un léger décalage sans doute lié à la différence de contenu ;

– le classement initial de `Google` est très bon ; le score du système aléatoire correspond à la proportion d'images pertinentes dans les 1000 réponses retournées et la précision est excellente en tête de liste, de l'ordre de 0.9, avec une précision moyenne autour de 0.73 ;

– les reclassements simples font sensiblement mieux que le hasard mais ne réussissent pas à faire mieux que le classement initial ;

– la fusion des reclassements simples fait mieux que le meilleur des deux mais toujours pas mieux que le classement initial ;

– la fusion entre le reclassement par le contenu visuel et le classement initial fait significativement mieux que le classement initial bien que le reclassement par le contenu visuel soit sensiblement moins bon que le classement initial ; cela arrive souvent quand les deux types d'informations fusionnées ont des sources très différentes, ce qui est le cas ici ;

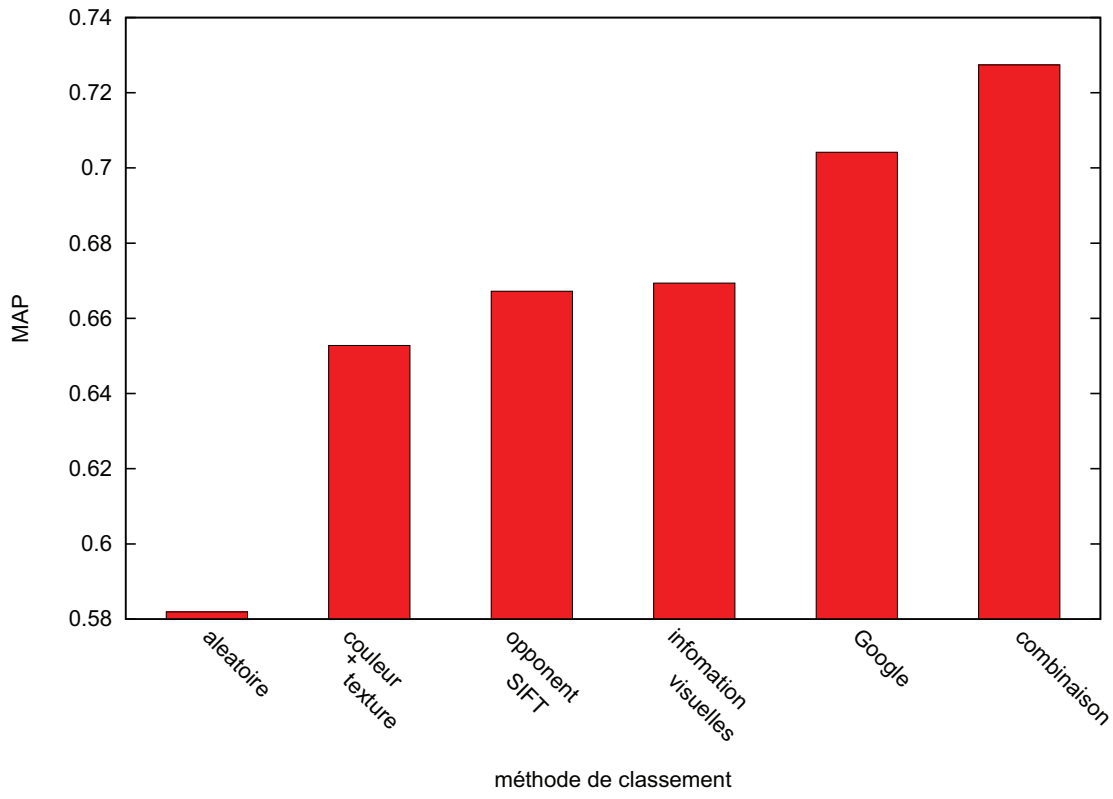

Figure 2. *Performances obtenues par les différentes méthodes sur l'ensemble de test*

– les optimisations faites sur l'ensemble de développement se sont révélées très robustes lors de la transposition sur l'ensemble de test ; le gain relatif en performance sur l'ensemble de test par rapport au classement initial est même légèrement supérieur sur l'ensemble de test (3.3% contre 2.6%) ;

– si l'on considère le gain sur le complément à 1 de la métrique MAP (chemin restant à parcourir pour obtenir un résultat parfait), il est de 6.7% sur l'ensemble de développement et de 7.8% sur l'ensemble de test.

La robustesse du réglage des paramètres entre l'ensemble de développement et l'ensemble de test est sans doute liée à la très bonne stabilité du système par rapport aux variations de ces paramètres autour de leurs valeurs optimales. Par exemple, dans le cas du nombre optimal de voisins pour le descripteur opponent SIFT, la figure 3 montre que les performances médianes sont peu affectées par le nombre de voisins considérés. Ceci explique sans doute aussi pourquoi aucun gain n'est obtenu avec une fonction f de fenêtrage. Les valeurs optimales trouvées pour le nombre optimal de voisins se situe entre 50 et 100 et la valeur optimale pour l'exposant α est proche de 0.4.

En ce qui concerne le temps de calcul, le plus long est de loin de calculer les descripteurs sur les images, ce temps est de quelques secondes par image mais les descripteurs peuvent avoir été calculés hors ligne sur toutes les images avant que les requêtes ne soient effectuées. La partie principale du calcul pour le reclassement est ensuite le calcul des plus proches voisins mais il peut être effectué en quelques cen-

taines de ms sur des listes de 1000 images avec les descripteurs considérés, ce qui est compatible avec une utilisation de la méthode comme post-traitement en ligne.

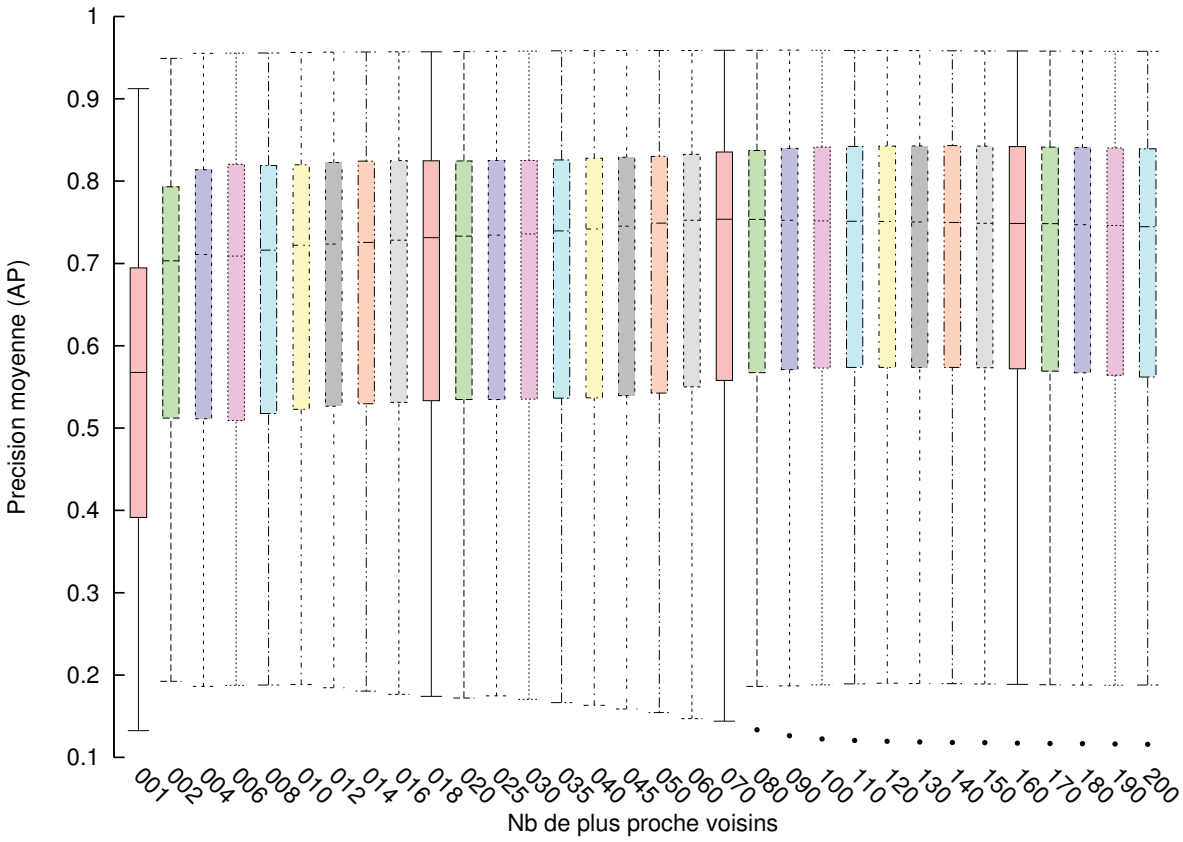

Figure 3. *Influence du nombre de voisins retenus sur la précision moyenne (AP), descripteur : opponent SIFT, distance : angle ; les boîtes montrent sur les 100 concepts de l'ensemble de développement, le minimum, le percentile à 25, la médiane, le percentile à 75 et le maximum.*

5. Conclusion

Nous avons présenté une méthode permettant de reclasser les images fournies par un moteur de recherche par mots-clés à l'échelle du web et à l'état de l'art. Cette méthode utilise le contenu visuel des images et elle est basée sur l'idée que les images pertinentes doivent être semblables entre elles et que les images non pertinentes doivent être différentes entre elle et des images pertinentes. Cette idée a été implémentée en classant les images en fonction de la distance moyenne de celles-ci avec leurs plus proches voisines. Le reclassement seul selon cette méthode ne fait pas mieux que le classement original du système de recherche mais, combiné à celui-ci, il permet un gain en performance relatif d'environ 3% en termes de précision moyenne et d'environ 7% si l'on considère le complément à 1 de la précision moyenne.

Ce gain est encore modeste mais statistiquement significatif et il est apporté à un système déjà très performant pour l'exploitation de l'ensemble des informations textuelles associées aux images. Plusieurs pistes peuvent être explorées pour améliorer

encore cette performance. Par exemple, une autre fonction que la fonction puissance peut être appliquée aux distances avant qu'elles ne soient moyennées. D'autres descripteurs peuvent aussi être utilisés à la place ou en complément de ceux déjà utilisés. La fusion au niveau des distances sur différents descripteurs (fusion intermédiaire) pourra également être considérée. Enfin, la méthode proposée pourrait être combinée avec une méthode favorisant la diversité des résultats retournés.

Remerciements

Ce travail a été réalisé partiellement dans le cadre du programme Quaero, financé par OSEO, l'agence française pour l'innovation. Nous remercions Winn Voravuthikunchai et ses collègues du laboratoire GREYC pour nous avoir fourni la collection de requêtes résolues qui nous ont permis d'évaluer notre approche.

6. Bibliographie

Bay H., Tuytelaars T., Van Gool L., « Surf : Speeded up robust features », *In ECCV*, vol. 1, p. 404-417, 2006.

Chechik G., Sharma V., Shalit U., Bengio S., « Large Scale Online Learning of Image Similarity Through Ranking », *Journal of Machine Learning Research*, vol. 11, p. 1109-1135, 2010.

Flickner M., Sawhney H., J. H., Q. D., B. G., Hafner M., Lee J., Petkovic D., Steel D., Yanker P., « Query by image and video content : The QBIC system », *IEEE Computer*, vol. 22, n° 12, p. 1349-1380, 1995.

Jégou H., Douze M., Schmid C., « Improving bag-of-features for large scale image search », *International Journal of Computer Vision*, vol. 87, n° 3, p. 316-336, feb, 2010.

Ke Y., Sukthankar R., « PCA-SIFT : a more distinctive representation for local image descriptors », *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, p. 506-513, 2004.

Lowe D. G., « Distinctive Image Features from Scale-Invariant Keypoints », *International Journal of Computer Vision*, vol. 60, n° 2, p. 91-110, 2004.

Mu Y., Yan S., Liu Y., Huang T., Zhou B., « Discriminative local binary patterns for human detection in personal album », *CVPR*, p. 1-8, 2008.

van de Sande K. E. A., Gevers T., Snoek C. G. M., « Evaluating Color Descriptors for Object and Scene Recognition », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, n° 9, p. 1582-1596, 2010.

Wengert C., Douze M., Jégou H., « Bag-of-colors for improved image search », *ACM Multimedia*, Scottsdale, United States, October, 2011. QUAERO.

Zhang D., Wong A., Indrawan M., Lu G., « Content-based Image Retrieval Using Gabor Texture Features », *IEEE Transactions PAMI*, p. 13-15, 2000.